How to Cite:
Arulmozhiselvi , S., & Indirani, G. (2022). Environmental air pollution monitoring in non industrial area using machine learning techniques and IOT. *International Journal of Health Sciences*, *6*(S7), 2115–2130. https://doi.org/10.53730/ijhs.v6nS7.11769

Environmental air pollution monitoring in non industrial area using machine learning techniques and IOT

S. Arulmozhiselvi
Research Scholar, Department of CSE, AU
Email: susips119@gmail.com

G. Indirani
Assistant Professor,Department of CSE Government college of Engineering Bargur, Tamil Nadu, India
Email: induk0992@gmail.com

***Abstract*---**Objectives: To provide an enhanced embedded-based IOT network for monitoring environmental pollution in non-Industrial areas with an efficient machine learning pollution prediction system. Methods: The methodology of the Dual processing Environmental Monitoring System (DPEMS) is carried out through a Dual processing unit (Arduino-Raspberry Pi) with advanced environmental air pollution, collecting sensors such as DHT22, CO2 (MG811), NO2 (MICS-4514), and SO2 (SGS-SO2). The environmental air pollutant data has been shared with IOT cloud storage from the Dual central processing unit to the IBM blue mix platform. To enhance a better pollution prediction system, machine learning classifiers such as ANN, SVM, and Decision Tree has been applied. The machine learning training and testing validation has been done using Pycharm 2021.1.1. The actual and predicted pollutant value has been evaluated using the performance metrics as RMSE, R2, and IA. Findings: the proposed IOT-based embedded DPEMS is utilized to increase the accuracy of real-time actual pollutant value and alert the threshold level of pollutant particles such as Temperature, Humidity, Carbon dioxide (CO2), and Nitrogen dioxide (NO2), and sulfur dioxide (SO2) in Non-Industrial areas. The enhanced three machine learning classifiers are implemented to improve their prediction accuracy values of Pollutant particles and alert the forecasting pollutant level. The actual and prediction accuracy of pollutant measurement has been increased by 5 % when compared to existing systems. The forecasting performance of DPEMS obtained the lowest average of RMSE of 31.5 with respect to pollutant gas concentration. Similarly, the highest average of RMSE of 35.6 has been predicted in the existing

International Journal of Health Sciences ISSN 2550-6978 E-ISSN 2550-696X © 2022.
Manuscript submitted: 9 April 2022, Manuscript revised: 18 June 2022, Accepted for publication: 27 July 2022

system (Arduino-based environmental monitoring system (AEMS)). The highest R2 and IA values of 0.85 and 0.93 were obtained from this proposed EMS system. Novelty/Application: the proposed EMS system is a more sensible model for collecting and monitoring actual pollutant levels when compared to existing systems (AEMS). The machine learning classifiers incorporated with IOT-based DPEMS have proved their better performance in forecasting pollutant prediction levels of non-Industrial areas and it more compatible model for collecting and forecasting pollutant levels in Industrial areas like the thermal power plant, chemical, and Pharma industries.

***Keywords*---**Environmental Monitoring, Internet of Things (IoT), Machine learning (ML), cloud computing, Embedded System, Artificial Neural Network (ANN), Support Vector Machine (SVM), DT (DECISION TREE).

Introduction

Nowadays, the rise in industrialization is sufficient to satisfy the requirements of the consistently developing population. The development of industrialization has created a ton of significant issues in the climate. As we are probably aware, many of the effects of air quality on the living creatures, humans, and environmental monitoring bodies of the world. It can cause sensitivity and harmful illnesses, such as cardiovascular infections and lung illnesses, as well as death.

The sensor data is transferred into the network via NB IoT devices to be processed in the cloud [1]. The Industries Exhaust particles make the significant air contamination due to the presence of carbon dioxides, nitrogen dioxides, Sulphur dioxides, temperature change, and humidity change. Using IoT technologies development [2-4]. The vast majority of the previous work in environmental air pollution monitoring has been tested by utilizing sampling and analysis methods. WSN-based pollution monitoring techniques are likewise experimented with, however, the real-time monitoring and wireless monitoring and alarming cycle request to increase the number of gateways for gathering information from all the nodes and require extra memory to predict the pollutant data. In this case, a machine learning-based pollutant prediction model with IoT enabled environmental monitoring system will be more suitable than existing techniques.

While different sorts of examination have been done in this expansive pollutant region, they are unable to state the quality of prediction in some cases. Salman Ahmad Siddiqui et al [7] prediction in high traffic area on a smart mirror using machine learning techniques such as ANN and Random Forest. Anabi Hilary Kelechi et al [8] The HTTP protocol has been utilized to send messages in bringing issues of poor air quality. [9] Ahmed Samy Moursi et al, proposed and evaluated a hybrid NARX architecture with machine learning algorithms for predicting PM2.5 concentration in the atmosphere. Broday et al [10] verify through a literature review how IoT device is used for building control (for energy-saving purposes) and monitor IEQ conditions inside buildings. From the literature survey, it was found that the existing method used a single board microprocessor system used to monitor pollution levels. Which has less efficiency due to high error rate.

So in this work, a mini-computer Raspberry Pi3-based IEMS has been planned and Integrated with IoT cloud service passage. Wi-Fi (ESP8266) has been utilized for data transmission between the microcontroller and cloud database. Thus, moving toward linear modeling techniques may not be appropriate for such information. In this project, we implemented advanced gas sensors and a dual-processor for the Industrial Environmental Monitoring system (IEMS) with IoT cloud service that will provide an effective solution to observe the real-time environment data, monitor and predict the Pollution level in Sethiyathope and kurinjipadi (non-Industrial regions) at Tamilnadu, India also real-time alerting system included. The air quality can be measured by using a parameter named Air Quality Index (AQI). Specifically, the aim of this work is to accurately predict concentrations of temperature, humidity, CO_2, NO_2, and SO_2 as they are considered to be the most predictable exhaust gas from the near industries plant at Kurinjipadi.

In our proposed work air pollution monitoring is a value addition for advanced gas sensors for predicting air pollution levels and using a dual processor with Real-time monitoring and alerting system and also enhanced with machine learning technics for best accuracy with Actual Environmental Monitoring System and Dual Prediction Environmental Monitoring System. Compared to an existing method, it gives the best result.

2. Methodology

The simplified hardware system architecture of the proposed system (IEMS) is shown in Figure.1. Raspberry pi is the mini arm computer that can control the environmental pollution parameters. The sensors are being used for detecting different environmental parameters present in a non-industrial area (Kurinjipadi) and Sethiyathope like temperature, humidity, Carbon Dioxide, Nitrogen Dioxide, and Sulfur dioxide. The sensors are connected to Arduino Board. The Arduino controller interfaced with the Raspberry pi 3 models through a USB cable. The data sensed by the environmental pollution sensors are continuously transmitted through Raspberry pi to the IoT cloud service (IBM Bluemix) by implementing the MQTT protocol.

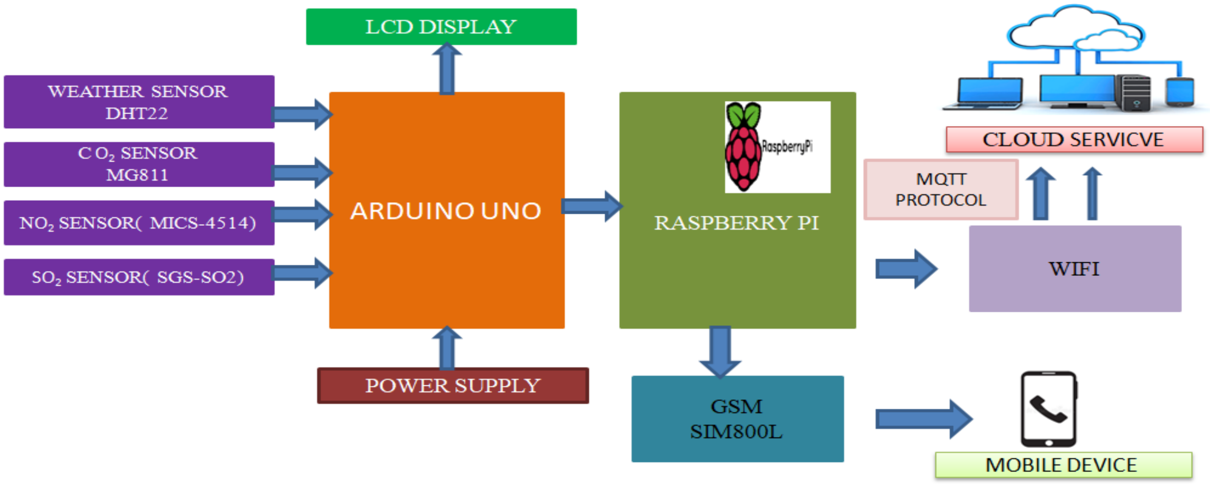


Fig.1 The Overall Hardware System Architecture of IEMS

Figure 1 depicts the overall hardware system architecture of the proposed system (IEMS). It incorporates four sensors that are connected to the Arduino controller interfaced with the Raspberry Pi single-board PC. The organization of network gateway through router has been implemented using MQTT protocol and Wi-Fi module to which hub is associated with the IBM Bluemix IoT platform

2.1 Sensing Unit:

In this proposed system, four sensors are used to measure the non-industrial environmental pollution parameters, namely the Weather sensor (DHT22), CO2 (MG811) Sensor, NO2 (MICS-4514) Sensor, SO2 Sensor (11-508). The weather sensor DHT22 is having digital outputs used for measuring temperature and humidity. The CO2 Sensor (MG811) has digital output used for measuring carbon dioxide. The NO2 (MICS-4514) is a 12-bit analog sensor and needs to be converted into digital by ADC. The DGS-SO2 sensor ID is a digital sensor connected to the digital input of the Arduino UNO controller. Arduino UNO is a low-cost, efficient microcontroller board based on the ATMEGA-328P family, which can be easily interfaced with Raspberry Pi and has a very effective ADC. Since Raspberry Pi 3 model collects the pollution data from Arduino controller and transmitted the processing data into the IoT platform (IBM Bluemix). The lightweight protocol MQTT (Message Queuing Telemetry Transport) is a protocol establishing communication between the raspberry pi and the IoT cloud server. The client can access the data that is being displayed on the dashboard by using the device id but the client will not be able to do any modification to the data received.

2.2 Processing Unit:

The Raspberry Pi 3 is furnished with a quad-core 64-bit Broadcom BCM2837 ARM Cortex-A53 SoC processor that can be running at 1.2 GHz, making it around half more impressive than the Raspberry Pi 2. And that implies the new

Raspberry Pi3 can be utilized for office applications and web browsing. The incredible advancement in Raspberry pi3 is without a doubt the expansion of a Wi-Fi chip and Bluetooth Low Energy. It additionally opens up more USB ports for interfacing the Arduino controller to collect environment pollution parameter data. Raspberry Pi has clarified that this new form is outfitted to the Internet of Things (IoT) and home computerization. The Raspberry Pi 3 is additionally viable with Windows 10 IoT Core, an operating system that has been installed for making and creating pollution monitoring dashboards in non-industrial areas. The Raspberry Pi 3 board is about the same size as the Raspberry Pi 2 and has a practically indistinguishable connector and part setup.

2.3 Data Integrated with IoT Cloud Service

In this proposed system, the open-source programming tool Node-RED has been used for IoT based environmental pollution monitoring. It is a highly used visual programming tool which helps to integrate Arduino and Raspberry pi 3. The Node red programming tool consists of thousands of flows and nodes that enable the user to connect all pollution parameter sensors. Flows can be run at the edge of the network on Raspberry pi 3. Node-Red provides a simple click mechanism to deploy the flows by the IoT hardware developers to a lightweight runtime industrial environment area. Here MQTT protocol is used to communicate the pollution monitoring data between Raspberry pi and IBM Bluemix cloud server. It is the mostly suited connectivity protocol for IoT Cloud service and minimize the additional interfacing device, low bandwidth, and unreliable network.

In Arduino section, the compiler has support ability to any programming language that has converted the program code to binary code. Arduino IDE Platform is used in proposed system to Received and process the pollution sensor output such as Temperature, Humidity, Co2, No2, and So2. As IDE is platform independent, it can run on Windows and Linux based operating system as well as Mac OS. A number of the key options of IDE embody a text console, message space, toolbar for common functions. A program for Arduino using IDE platform is known as sketch, Languages like C, C`++ are supported by Arduino IDE for programming.

2.4. Implementation and Prediction Model of Environmental Parameters

he implementation of non industrial environmental pollution monitoring system can be splitter into four models namely, measuring and sampling location, hardware setup model, software setup model and ML prediction model. The work flow diagram of proposed model is shown in figure 3.

2.4.1. Measuring and Sampling location

Two distinct time periods spanning seven days were selected to conduct the experiments at the two respective locations in Non-industrial area located in Kurinjipadi, and another is located in sethiyathope. The Non-industrial area consists of location as shown in fig 2. For the Non-industrial region, data were collected between 10 November and 2021and 10 December 2021. The measurement was taken at 1 -,8-, 12-, and 24-hour s from midnight. In total, the 5 environmental pollution parameters were measured as Temperature, Humidity, Carbon dioxide (CO2), Nitrogen dioxide (NO_2), and Sulphur dioxide (SO_2), In order to structure the prediction of pollution level in the proposed system the three-machine learning algorithm ANN, SVM, and DECISION TREE have been implemented.

Fig 2 Geomap of Sethiyathope and Kurinjipadi

In the above geographical map, the processed data from the time series data of one month has been analyzed by using machine learning algorithms such as ANN and SVM to predict the upcoming features of environment monitoring in choosing non-industries areas Kurinjipadi and Sethiyathope.

2.4.2. Hardware Setup model:

The IoT-based environmental pollution monitoring system measuring the pollution of two different regions is highly accurate, affordable, and easy to use. The digital sensors DHT22, CO2Sensor (MG811), and DGS-SO2 are connected to the digital input of the Arduino, whereas the sensor NO2 (**MICS-4514**) is connected to the Arduino controller through 12-bit ADC. An Arduino is interfaced with a Raspberry Pi 3 via a USB cable. The Raspberry Pi 3 is connected to the internet with the help of an inbuilt Wi-Fi chip. Initially, the operating system has to be installed on the Raspberry Pi by downloading an image from the Raspberry pi official website. Initially, the Arduino controller collected parameter sensing values from four pollution sensors and transmit the measured value to the Raspberry Pi controller. The Raspberry Pi controller is displayed by monitoring real-time pollution data on an LCD display. The measured value has been transmitted into the IBM Bluemix cloud service through the internet connection built into the WIFI module. Here, the communication between the Raspberry Pi and the IoT platform has been carried out by the MQTT protocol. In this proposed system, two alert message systems have been designed. One is the real-time monitoring data sent to the user through the GSM module connected to the Arduino controller, and the other is the alert message sent to the user through the IoT cloud service as shown in figure 4.

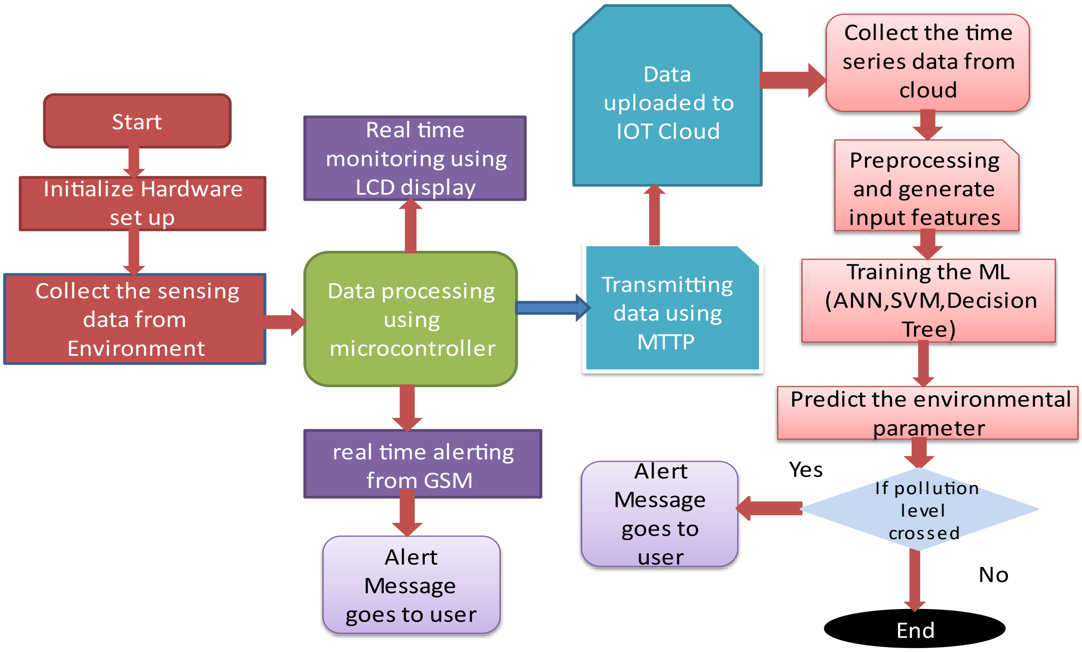


Fig 3 The Workflow Diagram of Proposed System (IEMS)

From the above flowchart, the user can view the real-time environmental data of industrial areas by logging into the IBM Bluemix cloud administration. The administrator can robotize certain methods such as auto prediction, real-time alerting, and cloud alerting service methods in this proposed system (EMS).

4.4.3. Software Setup Model

In order to use IBM's cloud services, an account is created at IBM Bluemix, and at the same time, the device has to be registered. Once the device is registered, the Bluemix IoT platform will acknowledge the user by providing the Authentication token which can be used for the communication of data from the device to the Bluemix IoT platform. The sensors are already connected with Arduino board and Raspberry pi for interfaced with Arduino. So, by deploying a flow containing the Serial in node to receive the data coming from the serial port on the Raspberry Pi, the Serial in a node is connected to the Watson IoT node for sending the data to the cloud storage. The cloud data can be seen on the dashboard of the IBM Bluemix IoT platform anywhere in the world, the only requirement is that the device should be connected to the internet.

4.5 Machine Learning Prediction Model:

The training and testing modules of the machine learning approach are shown in Fig. 4. It consists of data preprocessing, feature engineering, time series analysis, ANN prediction, SVM prediction, and random forest prediction model.

4.5.1 Data Pre-Processing:

The irrelevant features such as outliers and anomalies apart from the environmental pollution parameters data have been removed from the IoT cloud database. The training data set has been prepared to be a particular time series format for applying ML techniques like ANN, SVM, and RF. Data processing has been carried out in order to ensure the credibility of the machine learning prediction monitoring system.

4.5.2 Feature Engineering:

This step is concerned with selecting the features to be included in the prediction process along with each target gas, such as temperature, humidity, and day of the week. After collecting the pollution sensor values from the IoT cloud database, the target data is settled into 1,8,12, and 24 hour per day, weekday/weekend, and season by dividing it into tabulated format as per the need. ML prediction model can be applied to get the desired pollutant output.

4.5.3 Time Series Analysis:

This is a significant task with time series prediction of environmental pollutants in which a number of time series features for each input target are prepared in order to fulfil the analysis of time, date, week, month, hours, season, and environmental region. In target preparation, an average of the data points received over an hour is taken as the data which comes after 15 seconds, so there won't be a significant parameter change over such a short time and then the normalization is done on this data to avoid over-fitting.

4.5.4 ML Training Module:

In a prediction model, the machine learning architectures ANN, SVM, and DECISION TREE are trained and learn the assigned target feature in order to predict the unseen feature data based on historical data analysis from a cloud database. The process of enveloping the previous steps of the prediction model has been built and ML models are applied for predicting pollutant values of the future target as portrayed in Fig.1. In the ML training module, the measured environmental pollutant data is collected and the segmentation of selected features is applied to construct an ML prediction model on five pollution parameters. The various segment features selected the three ML algorithms, namely ANN, SVM, and DECISION TREE.

A.ANN Prediction:

ANN is the most well-known and broadly utilized supervised learning technique and requires a designer who develops a predicted result for any measured value. The feed-forward ANNs are frameworks of interconnected neurons that analyze the target data between input and hidden neurons, in which the associations have numeric loads that can be tuned in view of involvement. The neural structure of an ANN is shown in Fig 5.

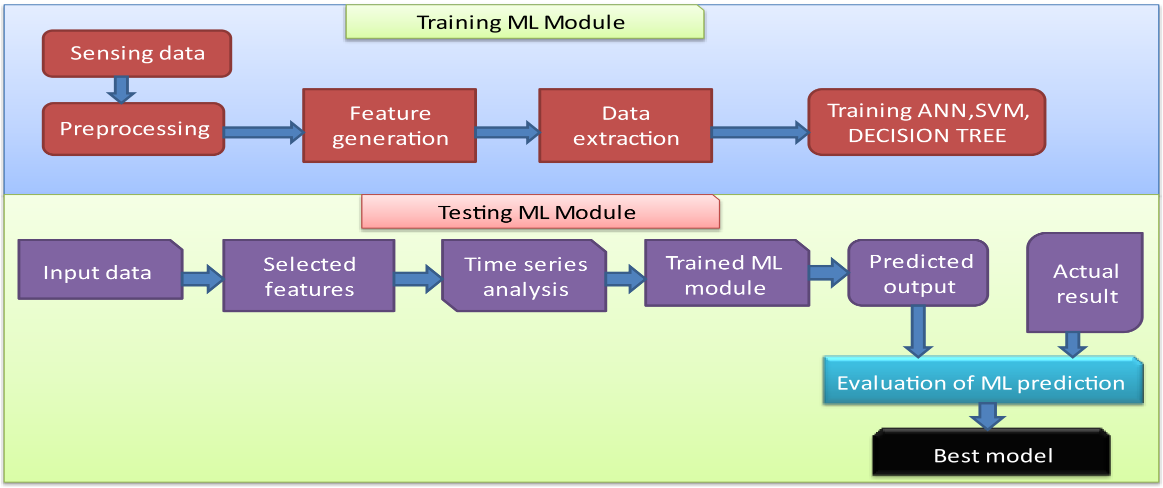


Fig 4 The Training and Testing Module of ML Techniques

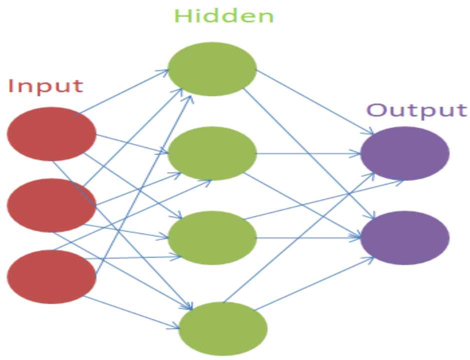


Fig 5 The Neural Structure of Feed Forward ANN

The training process of FFANN-BPs consists of two iterative steps, including the forward-propagation of the data stream and the back-propagation of the error signal. Firstly, original data is passed from the input layer to the output layer through the hidden processing layer. The input of the j^{th} neuron in the l^{th} layer $x^l_{jq}$is

$$x^l_{jq} = \sum_i \omega^l_{jiq}\, y^{i-1}_{iq} \qquad (1)$$

Whereω^l_{jiq} is the weight that connects the i-th neuron in the $(l–1)^{st}$ layer and the j^{th} neuron in the l^{th} layer, y^{i-1}_{iq}= f(x^l_{jq}) – θjqis the response of the j^{th} neuron in the l^{th} layer, and f is the activation function which is used to introduce the non-linearity into the network. If the predicted pollutant result is not matched consistently with the actual result, then the error signal is forwarded as feedback into the input neural network towards the direction of forwarding computing. The learning system comprises both forward and backpropagation. The ANN network powerfully looks through the weight which limits the network error in the weight space, arrives at the point of the memory cycle and the data extraction, and makes the predicted result of the network closer to the actual result.

B.SVM Prediction Model:

It is one of the most widely used supervised machine learning techniques for objective forecasting. The prediction model of the support vector attempts to give a non-linear mapping that is suitable for plotting the given environmental pollutant training data points, i.e., D: (x_1, y_1), $(x_2, y2)$..., (x_i, y_i) to a higher dimension, the support vector isolating decision boundary. A hyper-plane is one that recognizes the input actual data set over the maximum margin function.

In support vector regression, the basic idea is to map the original data x into a feature space F with high dimensionality by a non-linear mapping unknown function φ, and further to carry on linear regression in this space. Here, k (xi, $x_{j)}$ is named as the kernel function. Its value is equal to the inner product of two vectors x_i and x_j in the feature space (xi) and (x_j). So, the kernel function of the Map Reduction model is

$$H(x_i, x_j) = \varphi(x_i) * \varphi(x_j) \qquad (2)$$

The essential capacity of this SVM prediction helps to recognize a hyperplane in an N-layered hyperspace, which classifies the actual pollutant data points to recognize the predicted pollutant values in the middle of two classes. Our primary objective with the SVM Prediction Model is to observe a hyperplane that decides the number of classes that occurred in the cloud dataset, and the output of unsurveyed pollutant data will be predicted according to which class holds the most similarity with the new actual pollutant data. As for the regression problem, the non-linear function of the hyperplane has been constructed at the maximal margin with linear regression in this SVM prediction model.

C. Decision Tree prediction model

It is a general meta-approach to AI calculation that depends on choice tree learning calculation and bootstrap aggregation, which can be utilized for target prediction, classification, and different errands. The crucial and significant idea is to further develop the forecast precision where sub-tests assume a significant part. It is finished by fitting a large number of decision trees on random subsets to the feature accessible because of its abstinence from over-fitting. As for this situation, bootstrap aggregation is used (or called bagging) to persistently train decision or regression trees with arbitrary component subsets and test subsets over which the calculation is applied. The features for each bootstrap test are randomized. According to bagging theory, a random forest is a strong classifier based on multiple weak classifiers. Therefore, both the amount of data and the number of features of the subset are smaller than the original dataset. After all the trees are constructed, the unlabeled data are input into all decision trees. For each tree, $p(c_i)$is the estimated probability of the AQI levels i. The final probability of the AQI level $p'(c_i)$ in the random forest is defined in Equation (3), where T is the number of decision trees. the estimated probability $p'(c_i)$ can be measured by the equation 3.

$$p'(c_i) = \frac{1}{T}\sum_{k=1}^{T} p\,(c_i) \tag{3}$$

The final pollutant prediction result can be determined by Equation 4.

$$C'(i) = Max\big(p'(c_i)\big) \tag{4}$$

Where $C'(i)$ is the predicted class of AQI level.

4.6 Performance evaluation of ML

In order to assess the ML performance of the prediction model used and reveal any potential correlation between the predicted pollutant result and actual pollutant values, the following metrics are used in our experiments to evaluate the ANN, SVM, and DECISION TREE.

A. Root Mean Square Error (RMSE)

Root mean square error computes the square root of the mean for the square of the differences between predicted and actual values. The calculation of RMSE can be obtained using the eqn 5.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - A_i)^2}{n}} \tag{5}$$

Where n is the number of samples, Pi and Ai are the predicted and actual values, respectively. RMSE has the same measurement unit as the predicted or actual values, which in our study $\mu g/m3$. The lower the RMSE value, the better the model prediction performance.

B. **Coefficient of Determination (R^2)** This parameter evaluates the association between actual and predicted values mentioned in eqn 6.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(A_i - P_i)^2}{\sum_{i=1}^{n}(A_i - \bar{A}_i)^2} \tag{6}$$

Where n is the records count, Pi and Ai are the predicted and actual pollutant values, respectively. $\bar{A}$ represents the mean measured value of the pollutant.

C. Index of Agreement (IA)

A standardized measure is the degree of model forecasting error, varying between 0 and 1. This measure is described by Equation No 7.

$$IA = 1 - \frac{\sum_{i=1}^{n}(|\mathrm{Pi} - A_i|)^2}{\sum_{i=1}^{n}(|P -_i \bar{A}_i| + \quad |A_i - \bar{A}_i|)^2} \tag{7}$$

Where n is the samples count, Pi and Ai are the predicted and actual measurements, respectively. $\bar{P}$ and $\bar{A}$ represent the mean of predicted and measured value of the target, respectively

3. Result and Discussion:

In this proposed system, four sensors are used for measuring environmental pollution parameters such as temperature, humidity, CO_2, NO_2, and SO_2. All the sensors are connected to the Arduino Uno interfaced with the Raspberry Pi board. The existing system consists of an Arduino controller connected to DHT11, MQ6, MQ7, and MICS2714. In the proposed model, Arduino controllers collect sensing data from the environment and send it to the Raspberry board. The Raspberry Pi 3 model B board can operate in two modes: one is the data acquisition mode and the other one is the web server mode. The Raspberry Pi 3 communicates with the IBM Bluemix IoT platform via the MQTT protocol to send sensing data. The transmitted data will be displayed on an LCD display, and the cloud service will provide a pollution alert message to the user. The training and testing of machine learning techniques have been carried out by Python 3.8 with PyCharm community edition 2021.1.1.

Environmental pollution concentrations have been measured in non-industrial areas. Initially, the environmental pollution monitoring has been done in Kurinjipadi and Sethiyathope. The readings were taken at every 4 hours interval during the day. Environmental pollution prediction has been carried out by applying three machine learning techniques, namely ANN, SVM, and Random Forest. Machine learning processing is the prediction methodology of environmental data points from IOT cloud database storage. The forecasting prediction models' performances are measured using RMSE, R2, and I2. Moreover, the results are visualized for easy comparisons and testing of the output from the different models.

Table 1: The actual pollutant values are obtained by AEMS and DPEMS

Hours	Temp°c		Humidity (%)		CO_2 (PPM)		NO_2 (PPM)		SO_2 (PPM)	
	ACTAEMS	**ACTDPEMS**	**ACTAEMS**	**ACTDPEMS**	**ACTAEMS**	**ACTDPEMS**	**ACTAEMS**	**ACTDPEMS**	**ACTAEMS**	**ACTDPEMS**
1	16.1	19.5	28.6	32.5	3.5	5	3.6	6.5	3.8	4.2
8	22.8	27.5	24	28	4.5	6.5	4.3	6.5	2.5	4.1
12	16.4	22.1	21	24	9.3	12.5	9.8	12	4.5	7.5
24	18.7	22.7	29	34	5.1	7.5	7.9	10	3.8	6.2

In table 1 the comparative analysis of DPEMS and AEMS to measure the actual pollutant value for each parameter is tabulated. Real-time environmental pollution parameters are measured and monitored with AEMS and DPEMS with respect to 1, 8, 12, and 24 hours every day. Table 1 shows that the obtained actual environmental pollutant values of the proposed system (DPEMS) are increased by 5%, 6.5%, 4.5%, 5.5%, and 6.5% in terms of temperature, humidity, CO2, NO2, and SO2 when compared to the existing system.

Table 2: The Prediction Variance of AEMS and DPEMS

Hours	Temp°c		Humidity (%)		CO_2 (PPM)		NO_2 (PPM)		SO_2 (PPM)	
	PVAEMS	**PVDPEMS**	**PVAEMS**	**PVDPEMS**	**PVAEMS**	**PVDPEMS**	**PVAEMS**	**PVDPEMS**	**PVAEMS**	**PVDPEMS**
1	2.8	1.8	3.1	1.65	2.6	0.98	5.1	2.1	4.2	1.2
8	3.5	1.5	2.95	0.98	3.5	1.5	2.6	1.2	3.9	0.98
12	2.2	0.95	6.5	2.4	4.2	1.6	5.1	0.98	2.8	0.88
24	4.2	0.75	4.5	3.1	3.9	1.8	3.6	0.92	1.9	0.9

The comparative analysis of AEMS and DPEMS in terms of prediction variance of each pollutant parameter is tabulated in table 2. Table 2 shows that the lowest prediction variance values of 0.95, 2.4, 1.6, 0.98, and 0.88 are obtained for each pollutant parameter at the 12th hour of each day. It is observed from table 2 that the pollutant prediction accuracy of the proposed system against the existing system has been improved by 3.5 %, 3.8 %, 4.3%, 5.1%, and 3.8% for each pollutant parameter, respectively.

Table 3: The Performance Evaluation Result of Machine Learning Techniques

Parameter	ANN			SVM			DECISION TREE		
	RMSE	**R^2**	IA	**RMSE**	**R^2**	IA	**RMSE**	**R^2**	IA
Temp(°C)	43.5	0.54	0.85	31.5	0.65	0.91	21.5	0.54	0.86
Humidity (%)	22.5	0.67	0.79	27.6	0.67	0.83	34.1	0.62	0.72
CO_2(PPM)	31.6	0.63	0.75	32.5	0.74	0.95	41.5	0.52	0.92
NO_2(PPM)	35.6	0.47	0.91	37.4	0.52	0.86	31.1	0.68	0.94
SO_2(PPM)	15.5	0.57	0.85	17.5	0.64	0.94	28.5	0.55	0.88

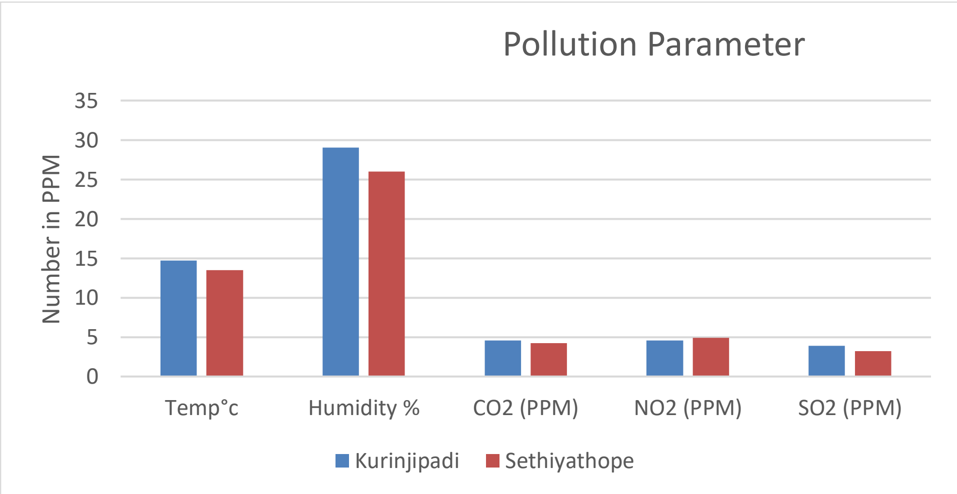
The machine learning performance of the prediction level has been evaluated with performance metrics such as RMSE, R2, and IA, which are calculated in table 3. As per table 4, the lowest average RMSE error value of 31.5 % was obtained by the proposed system. Similarly, the highest average RMSE value of 35.6% was obtained by the existing system. Overall, the RMSE error value has been reduced by 4.1 % that was done by the proposed model. Figures 5 and 6 show that the highest R2 value of 0.85 was obtained by DPEMS, whereas the lowest R2 of 0.47

was exhibited by AEMS. Similarly, the highest and lowest IA values of 0.92 and 0.85 have been obtained by DPEMS and AEMS, respectively.

Table 4: The Obtained Performance Metrics Result of ML Techniques

ML Algorithm	RMSE	R2	IA
ANN	31.8	0.68	0.93
SVM	27.56	0.71	0.96
DECISION TREE	32.81	0.53	0.88

The average performance metrics results for the three machine learning techniques are calculated and tabulated in table 4. To predict the environmental pollutants parameters, the classifier SVM from table 4 is combined with the DPEMS best forecasting model. The comparison of environmental pollutant prediction values for each parameter in Kurinjipadi and Sethiyathope are shown in Figure 6.



Note: Location 1- Kurinjipadi , Location 2 – Sethiyathope
Figure .6 the average prediction value of pollutant parameters in Non Industrial Region

Conclusion

In this work, the enhanced Dual Processing Environmental Monitoring System (DPEMS) is experimented with by implementing a dual hardware setup as an (Arduino-Raspberry Pi) controller and advanced pollution monitoring sensors. The proposed system has been enhanced in order to collect actual pollutant parameter values from an Arduino-based environmental monitoring system (AEMS). The sensing data is transmitted into the IOT cloud service (IBM Bluemix platform) efficiently by using the MQTT protocol. Every day, the sensing data of environmental pollutant parameters are monitored at 1, 8, and 24 hours after midnight. The proposed system (DPEMS) has achieved against the existing system to measure actual pollutants' accuracy values increased by 5%, 6%, 4.5%, 5.5%, and 6.5% with respect to temperature, humidity, CO_2, NO_2, and SO_2. Similarly, the DPEMS combined with a machine learning system increased the accuracy of pollutant prediction by 3.5%, 3.8%, 4.3%, 5.1%, and 3.8%, respectively. The highest average of RMSE values was 31.5, obtained in the proposed system (DPEMS) with machine learning pollutant prediction. Whereas the existing system has an RMSE of 35.6 for pollutant forecasting, this proposed DPEMS has the highest average of R^2 and IA at 0.85 and 0.93, respectively. In addition to the machine learning performance analysis, SVM outperforms other algorithms for the prediction of environmental pollution prediction levels in terms of RMSE, R^2, and IA. On the other hand, Decision Tree achieved the worst results because of its poor generalization ability when working on a small dataset with many attributes, leading to a complex network that over fit the data, while having ANN better than Decision Tree in our case due to its adaptability with high-dimensional data. In future work, we can experiment with the proposed DPEMS system in highly polluted areas like thermal power plants, chemical and pharma industries where wireless sensors are not suitable for forecasting environmental pollution levels.

Reference

[1] IoT enabled Environmental Air Pollution Monitoring and Rerouting system using Machine learning algorithms Leeban Moses, Tamilselvan, Raju, Karthikeyan.IOP Conf. Series: Materials Science and Engineering 955 (2020) 012005 IOP Publishing doi:10.1088/1757-899X/955/1/012005

[2] Yu-Lin Zhao , Jiali Tang , Han-Pang Huang , Ze Wang , Tse-Lun Chen , Chih-Wei Chiang5 , Pen-Chi Chiang," Development of IoT Technologies for Air Pollution Prevention and Improvement" in : Taiwan Association for Aerosol Research Available from: https://doi.org/10.4209/aaqr.2020.05.0255

[3] Y. Qi, Q. Li, H. Karimian, and D. Liu, "A hybrid model for spatIoT Temporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory," Science of The Total Environment, vol. 664, pp. 1–10, May 2019, Available from:https://doi.org/10.1016/j.scitotenv.2019.01.333.

[4] J. Ma, Z. Yu, Y. Qu, J. Xu, and Y. Cao, "Application of the XGBoost Machine Learning Method in PM2.5 Prediction: A Case Study of Shanghai," Aerosol and Air Quality Research, vol. 20, no. 1, pp. 128–138, 2020, Available from: https://doi.org/10.4209/aaqr.2019.08.0408.

[5] JunHo Jo , 1 ByungWan Jo , 1 JungHoon Kim , 1 SungJun Kim,1 and WoonYong Han2 "Development of an IoT-Based Indoor Air Quality Monitoring Platform" Vol 2020. https://doi.org/10.1155/2020/8749764.

[6] Cristian Toma, Andrei Alexandru, Marius Popa,and Alin Zamfiroiu"IoT Solution for Smart Cities Pollution Monitoring and Security challenges" 2019 Aug; 19(15): 3401.https://dx.doi.org/10.3390%2Fs19153401.

[7] Salman Ahmad Siddiqui*, Neda Fatima, Anwar Ahmad," Smart Air Pollution Monitoring System with Smog Prediction Model using Machine Learning ".") International Journal of Advanced Computer Science and Applications", Vol. 12, No. 8, 2021 Available from:https://doi.org/10.14569/IJACSA.2021.0120846

[8] Anabi Hilary Kelechi1 , Mohammed H. Alsharif2 , Chidumebi Agbaetuo3 ," Design of a Low-Cost Air Quality Monitoring System Using Arduino and Thing Speak ""techscience press computer materials and continua ,2021 Available from:https://doi.org/10.32604/cmc.2022.019431

[9] Moursi, A.S., El-Fishawy, N., Djahel, S. et al. (2021). An IoT enabled system for enhanced air quality monitoring and prediction on the edge. Complex and Intelligent Systems. 7. 2923–2947 Available from: https://doi.org/10.1007/s40747-021-00476

[10] Broday EE, da Silva MC. The role of internet of things (IoT) in the assessment and communication of indoor environmental quality (IEQ) in buildings: a review. Smart and Sustainable Built Environment. 2022 Mar 1. Available from: https://doi.org/10.1016/j.envres.2021.112574

[11] IoT-EMS: An Internet of Things Based Environment Monitoring System in Volunteer Computing Environment Sourav Kumar Bhoi , Sanjaya Kumar Panda , Kalyan Kumar Jena , Kshira Sagar Sahoo , N. Z. Jhanjhi , Mehedi Masud and Sultan Aljahdali.Intelligent Automation & Soft Computing Available from:https://doi.org/ 10.32604/iasc.2022.022833

[12] PranveraKortoçi, Naser Hossein Motlagh, Martha ArbayaniZaidan, Pak Lun Fung. et al. (2022). Air pollution exposure monitoring using portable low-cost air quality sensors. Smart Health. 23. 100241. https://doi.org/10.1016/j.smhl.2021.100241

[13] Saad Hikmat Haji and Amira B. Sallow, “IoT for Smart Environment Monitoring Based on Python: A Review, AJRCOS. May 2021 Available fromhttp://dx.doi.org/10.9734/ajrcos/2021/v9i130215.

[14] Mario J. Diván , Senior Member, IEEE, María Laura Sánchez-Reynoso , Juan Esteban Panebianco, and Mariano J. Méndez”IoT-Based Approaches for Monitoring the Particulate Matter and Its Impact on Health” IEEE INTERNET OF THINGS JOURNAL, VOL. 8, NO. 15, AUGUST 1, 202I. https://doi.org/10.1109/JIOT.2021.3068898

[15] Ali AbaasAboodee Al-Zaheiree, Yasir Ahmed Taha AL-Zubaidi, S.S.Gaikwad, R.K.Kamat “Advanced Air Pollution Detection using IOT and Raspberry PI” International Journal of Recent Technology and Engineering (IJRTE), Vol-8 January 2020. Available from: DOI:10.35940/ijrte.E5931.018520.

[16] Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2021). Get vaccinated when it is your turn and follow the local guidelines. *International Journal of Health Sciences*, *5*(3), x-xv. https://doi.org/10.53730/ijhs.v5n3.2938